Request for Information: NNH15ZDA012L
Respondent Name: K. Winslow Farrell, Jr.
Institution: Near Real Time Corporation
E-Mail address: winslowfarrell@yahoo.com

Topic Addressed (Item Number 7, regarding search and improved access to data):

Retrieval of source data from the Planetary Data System (PDS) that precisely resolves time-critical scientific questions can be cost-effectively performed through stored, pre-specified *"persistent stare"* search queries that retrieve information conforming to a consistently applied, standardized data management format.

Rationale – The role of information assurance in scientific data validation:

In two NASA-sponsored workshops held in 2015 for landing site selection of robotic and human missions to Mars, researchers presented scientific investigations, recommendations and conclusions of the merits of different Martian locations. In these workshops, presentations were reviewed in real-time against a rubric of pre-determined scientific questions. Workshop audience members judged the effectiveness of arguments made in presentations and the degree of match of Martian locations against pre-specified scientific criteria. Each presentation attempted to provide persuasive arguments on the degree of fit of a specific Martian location against pre-determined requirements that encompassed a variety of scientific questions of importance to the planetary community. The relative value of a location relied on the scientific interpretation of the location's merits, hazards, and geomorphology, which in turn was based on the selection, reduction and normalization of a large volume of sensor information ranging across visible and infrared sensor data taken from Martian orbit, radar traces, in-situ data, terrestrial analogs, and referenced material. Taking into account an entire workshop's presented information, scientific evidence was based on a large number of electronic data products retrieved from the PDS, which formed the underlying fact base of each presentation. However, this primary data was not directly available to the workshop attendees. The audience of researchers who reviewed these presentations could not easily retrieve the original raw data or underlying algorithms used to process the data into the assertions presented. Consequently, recommendations and polled votes supporting sites were made on the basis of subjective judgments that lacked the ability to directly inspect the electronic data products from PDS upon which those judgments relied.

The importance of assured end-to-end data integrity from sensor to final report is widespread beyond the issues faced in NASA landing site workshops. The United States Air Force (USAF) currently seeks innovative solutions to the issue of multi-sensor fusion for target identification in quickly evolving battlespace environments. Across USAF platforms, processing chains that transform raw sensor data to actionable information has largely remained sensor-dependent. However, the pace of the Observe, Orient, Decide, and Act (OODA) decision-making loop in USAF

environments has intensified; the time-critical nature of target identification has forced the information assurance and integrity of data elements to be traced from original raw sensor signatures, through transformations in on-board and off-board algorithms, across databases to final evidence, normalized to alerts and human readable, accessible visualizations. The resolution to both the USAF battlespace targeting decision and NASA planetary landing site location decision could be achieved through a set of operationally proven information assurance techniques in critical multi-sensor environments.

An approach to address the data integrity required for NASA and USAF decision-making regimes includes pre-specification of electronic data that corresponds to the precise questions addressed. In response to the 9.11 attacks, the United States Intelligence Community (IC) endeavored to create a responsive means to reduce large quantities of disparate information, collected through a variety of means into relevant, timely and accurate identifiers of the patterns of behavior of individual humans and organizations. Across agencies, analysts found that they could improve their efficiency in search for relevant data through the creation of search terms that would match new data sources as well as historic data resident in repositories. The ability to create these pre-defined search terms, including complex terminology that was highly domain-specific and reliant on the signatures of identifiers in specific sensors, was made more effective once pre-defined queries were constructed. In real-time, pre-built queries filtered high volume data streams, allowing processing of those filtered results into information that was presented to human analysts as alerts. These "*persistent stare*" queries, formulated in a high-level query language, allowed collaboration among IC analysts to retrieve precise matches to queries for events, patterns and behaviors identified in sensed data across a variety of locations and circumstances. These queries were pre-built, stored and constantly tuned to retrieve up-to-date, as well as historically relevant data for analytic investigations. Moreover, the queries themselves and their corresponding results were retained to allow other researchers to have consistent access to source documents, thereby assuring information integrity and so-called data provenance – defined as the ability to trace data from their origin, across processing environments, to their use as a rapidly available fact base.

These three scenarios, involving time-critical decisions for NASA, USAF and the IC, illustrate the importance of having PDS data available to scientific researchers as a readily available fact base, suitable to direct scrutiny, independent audit, and peer review. The importance of data provenance management in scientific data repositories ranges well beyond the comparison of Martian landscapes as suitable locations for scientific investigation or resource location for human habitat. The wider lens of USAF multi-sensor fusion in a time-constrained environment illuminates the comprehensive value of access to reliable information. Data that corresponds to pre-defined, transparent query criteria brings a contingent relevance perspective to PDS electronic data products. Currently unremarkable PDS data could be critical to future judgments and inquires. For example, 1975 era biological data from the Viking 1 lander has been re-examined and reinterpreted

after decades in data storage. PDS Products that are tied to precise queries and retained in a query-results data structure could be used for comparative planetology, as a corpus of responses to search-term based entities that remain relevant to the evolving trajectory of scientific inquiry and mission planning.

Suggested Improvements to the Current PDS Search Capabilities:

Planetary researchers could improve efficiency in search capabilities for planetary data in the PDS through the construction, maintenance and tuning of *Persistent Stare* queries, corresponding to current, historical, contingent or projected questions of scientific importance, in a language formulated to access data from the PDS repository. These questions could correspond to an investigative rubric vetted across the scientific community. Semantics could be debated and resolved in advance of time-critical decisions, so that results that precisely match semantically defined queries could serve as guideposts for investigations and decisions. Ideally, investigators themselves would create these queries, as a part of peer-reviewed research they would conduct, formulated in a high-level query language, such as Pattern Specification Language (PSL), a standard in use throughout the United States Intelligence Community (IC). These queries would be pre-built, retained as an appendix for further review, and matched with electronic data that gave rise to evidence upon which scientific conclusions were based. The PDS would retain a record of these queries, allowing retrieval for in-depth investigations, and retained to allow other researchers to have consistent access to source documents, so that scientific research could be rapidly reproduced and extended, rather than repeated. As a aggregate resource, the queries themselves would provide a consistent, normalized and articulated corpus of detailed investigation, and therefore provide consistent data provenance to the research community. A robust PDS data provenance management system would allow data to be traced from original raw sensor signatures, to transformations in databases, to final scientific evidence through a set of scientifically validated data normalization and data fusion techniques across both single sensor studies and multi-sensor environments.

Discussion:
*Possible Impact of **not** Implementing Suggested "Persistent Stare" Queries*

Currently, planetary data associated with scientific investigations may not adhere to principles of information assurance -- managing risks associated with maintaining availability and integrity of authentic source data in the PDS that corresponds to information upon which scientific assertions are made. With current information assurance practices in the PDS, reviewers may not be able to rapidly reproduce critical scientific results, particularly in time sensitive circumstances, as in landing site selection. Maintaining the current PDS approach to information assurance may limit the availability of an underlying fact base -- decisions of significance may be undertaken while data is unavailable for confirmation. Consequently, relevant PDS electronic data may only directly match investigative queries conducted during the time of an original investigation. Without a means to rapidly retrieve salient queries

and their commensurate results, the range of assertions, assumptions and conclusions could be limited, leading to ephemeral recommendations that lose relevance over time.

Potential Impacts of Implementing Suggested "Persistent Stare" Queries

We can expect positive and negative impacts associated with the costs and benefits of building a corpus of queries that correspond to the management of results that match scientific investigations, in keeping with modeled lessons learned from implementing *"persistent stare"* queries within the United States Intelligence Community (IC). These lessons include the development of query standards that match the fidelity of the data structures of electronic data at the PDS, adherence to those standards in the creation of accurately implemented queries by investigators, training investigators to create appropriate queries, and managing the evolution of those queries to encompass both historic and unanticipated future sensor data streams. If the planetary scientific community implemented a similar query corpus associated with current and anticipated scientific inquiries in the PDS, we could anticipate a loss of speculation, and a degree of transparency of scientific inquiry that could improve the accuracy, timeliness and integrity of data-driven conclusions.

K. Winslow Farrell, Jr. was a member of the Viking Flight Team and Landing Site Certification Team. His consultancy, Near Real Time Corporation, has provided advice to aerospace, technology and service companies, and agencies in the United States government since 2000.